



## Does accounting for imperfect detection improve species distribution models?

Christopher T. Rota, Robert J. Fletcher, Jr, Jason M. Evans and Richard L. Hutto

C. T. Rota (*ctr4g2@mail.missouri.edu*), R. J. Fletcher, Jr and J. M. Evans, *Univ. Florida, Wildlife Ecology and Conservation, PO Box 110430, Gainesville, FL 32611-0430, USA*. – R. L. Hutto, *Univ. of Montana, Div. of Biological Sciences, Missoula, MT 59812, USA*.

Models of species distributions are increasingly being used to address a variety of problems in conservation biology. In many applications, perfect or constant detectability of species, given presence, is assumed. While this problem has been acknowledged and addressed through the development of occupancy models, we still know little regarding whether addressing the potential for imperfect detection improves the predictive performance of species distribution models in nature. Here, we contrast logistic regression models of species occurrence that do not correct for detectability to hierarchical occupancy models that explicitly estimate and adjust for detectability, and maximum entropy models that attempt to circumvent the detectability problem by using data from known presence locations only. We use a large-scale, long-term monitoring database across western Montana and northern Idaho to contrast these models for nine landbird species that cover a broad spectrum in detectability. Overall, occupancy models were similar to or better than other approaches in terms of predictive accuracy, as measured by the Area Under the ROC Curve (AUC) and Kappa, with maximum entropy tending to provide the lowest predictive accuracy. Models varied in the types of errors associated with predictions, such that some model approaches may be preferred over others in certain situations. As expected, predictive performance varied across a gradient in species detectability, with logistic regression providing lower relative performance for less detectable species and Maxent providing lower performance for highly detectable species. We conclude by discussing the advantages and limitations to each approach for developing large-scale species distribution models.

Accurate predictions of species distributions are essential for conservation (Carroll and Johnson 2008), management (Fernandez et al. 2006), and forecasting the biological effects of global change (Lassalle et al. 2008). Species distribution models (SDMs) are a commonly used tool to describe the geographic distribution of plants and animals by quantifying and extrapolating species–environment relationships (Guisan and Thuiller 2005). Recent advances in statistical techniques and geographic information system (GIS) technology has resulted in a proliferation in the number of SDM approaches (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Elith et al. 2006). While applications exist for predicting both the abundance (Thogmartin et al. 2004) and occurrence (Guisan and Zimmermann 2000) of species across space, often only information on occurrence is available for developing models.

Logistic regression is a commonly used approach for modeling species occurrence (Guisan and Zimmermann 2000). This approach relies on binomial or binary response data in the form of presences and absences (or more frequently, detection/non-detection) (Manly et al. 2002). Obtaining reliable presence/absence data is often costly and difficult. Additionally, records of occurrence of many

species are sparse or known only through museum records. This has prompted the development of a variety of techniques for modeling presence-only data. A recent comprehensive analysis of presence-only SDMs demonstrates that many of these techniques have high predictive performance (Elith et al. 2006).

A common problem with the above approaches is their inability to properly account for the imperfect detection of species. For example, SDM approaches that rely on presence/absence data, such as logistic regression, require the strong assumption that detection probability is constant across sites – an assumption that is unlikely to hold in most situations (MacKenzie et al. 2002). Further, recent investigations have demonstrated that failure to account for imperfect detection can lead to biased estimates of habitat relationships for logistic regression models (Tyre et al. 2003, Gu and Swihart 2004, Martin et al. 2005). Presence-only methods attempt to circumvent the detectability issue by only using data from locations where a species was known to occur (Brotons et al. 2004). As a result, however, presence-only methods are unable to directly estimate the probability of occurrence (Ward et al. 2008), but instead estimate a relative measure of the probability of occurrence (Phillips et al. 2006). Additionally, presence-only methods

require a random (i.e. unbiased) sample of points from the landscape. However, since many presence-only SDM approaches rely on museum specimens or incidental observations of rare species, this assumption is rarely met (Phillips et al. 2009).

Recently developed occupancy modeling approaches that account for imperfect detection (hereafter, occupancy modeling approaches) have the potential to overcome this shortfall by simultaneously estimating the probability of occurrence and the probability a species is detected (MacKenzie et al. 2002, 2006). These approaches rely on detection/non-detection data and require multiple surveys at a site to estimate the probability of detection (MacKenzie et al. 2002). Occupancy models can accommodate covariates associated with detection probability and the probability of occurrence, and can be estimated with both a likelihood-based (MacKenzie et al. 2006) and a hierarchical Bayesian approach (Royle and Kéry 2007, Royle and Dorazio 2008).

Occupancy modeling approaches have gained enormously in popularity (Marsh and Trenham 2008) and have the potential to improve the predictive performance of SDMs. However, to our knowledge they have yet to be assessed relative to other approaches in a species distribution modeling context, where the goal is to maximize the predictive performance of models across geographic regions. Here, we use occupancy modeling approaches to create SDMs for breeding forest birds that are imperfectly detected. We contrast an occupancy modeling approach with two other commonly used SDM approaches: logistic regression and maximum entropy, a presence-only approach that performs well compared to other presence-only approaches (Elith et al. 2006, Phillips et al. 2006). By explicitly accounting for imperfect detection when building SDMs, we expect that occupancy models should produce more accurate estimates of habitat relationships and improve predictive performance relative to approaches that do not account for detection bias (e.g. logistic regression) or use a subset of the data to do so (e.g. presence-only approaches). We further expect that occupancy models should provide greater improvements in predictive performance for difficult-to-detect species relative to highly detectable species. We test these expectations using an extensive database on bird distributions that uniquely accounts for species detectability across a broad spatial (Montana and Idaho, USA) and temporal (1994–2008) range.

## Methods

### Modeling database

Our analysis draws on a spatially and temporally extensive effort to estimate geographical distributions and monitor population trends of landbirds breeding in western Montana and northern Idaho. Since 1994, surveys coordinated through the Northern Region Landbird Monitoring Program (NRLMP) have monitored all diurnal landbirds that can be detected using a single point-count methodology (Hutto and Young 1999). The entire NRLMP dataset consists of 482 permanently marked transects stratified

across lands administered by nine National Forests and three cooperating agencies (Fig. 1).

Each transect consists of 10 permanently marked points (sites hereafter), spaced ca 300 m apart. Standard 10-min, 100-m radius point-count surveys were conducted at each permanently marked site. During each survey, all birds seen or heard were recorded by trained observers. A subset of transects were surveyed every year from 1994 to 2008, with the exception of 1997, 1999, 2001, and 2006. Each site was only surveyed once per breeding season (Hutto and Young 2002). During all years, transects were surveyed from late-May to mid-July.

In many years, each 10-min survey was divided into two 5-min sampling intervals. When point-counts were divided into shorter sampling intervals, observers noted the sampling interval in which a species was first detected. By doing so, observers were effectively employing a “removal” sampling design (sensu MacKenzie and Royle 2005, Rota et al. 2009), meaning they were only sampling for a species until it was first detected. We used this information to estimate detectability in SDMs (see below).

### Transect and species selection

We initially refined our analysis to transects and sites with reliable GIS data. Occupancy modeling approaches require multiple surveys at each site to estimate detection probability. Thus, to build SDMs that incorporate imperfect detection, we further refined our analysis to those sites where surveys were subdivided into shorter sampling intervals (i.e. two 5-min intervals). Finally, we removed from the analysis surveys conducted during 2003 and 2005 because of restricted geographic sampling. This resulted in 8035 point-count surveys, located on 268 transects (Fig. 1), spanning 5 different years (1994, 1995, 2004, 2007, and 2008). We then split the refined data into two groups: training and validation data. We randomly selected 2/3 of the transects (5745 point-counts located on 192 transects) for model training, i.e. we developed SDMs using these data only. The remaining 1/3 of the data (2290 point-counts located on 76 transects) was used to validate SDMs.

For this analysis, we chose species along a gradient of estimated detection probability. To ensure enough data for developing SDMs, we limited our pool of candidate species to those detected on at least 5% of point-counts, for a total of 39 species. We then fit intercept-only occupancy models (MacKenzie et al. 2002) to each of these species to estimate relative detectability. Estimates of detection probability,  $p$ , for a 5-min sampling interval ranged from 0.29 to 0.81. We used these estimates to categorize species as “difficult to detect” ( $p=0.29$  to 0.49), “moderately detectable” ( $p=0.50$ –0.74), and “highly detectable” ( $p=0.75$ –0.81), and selected three birds from each group for which adequate habitat layers were available. Specifically, we selected pileated woodpecker, brown-headed cowbird, and Townsend’s solitaire from the “difficult to detect” category; chestnut-backed chickadee, golden-crowned kinglet, and MacGillivray’s warbler from the “moderately detectable” category; and Townsend’s warbler, varied thrush, and Swainson’s thrush from the “highly detectable” category (see Table 1 for scientific names).

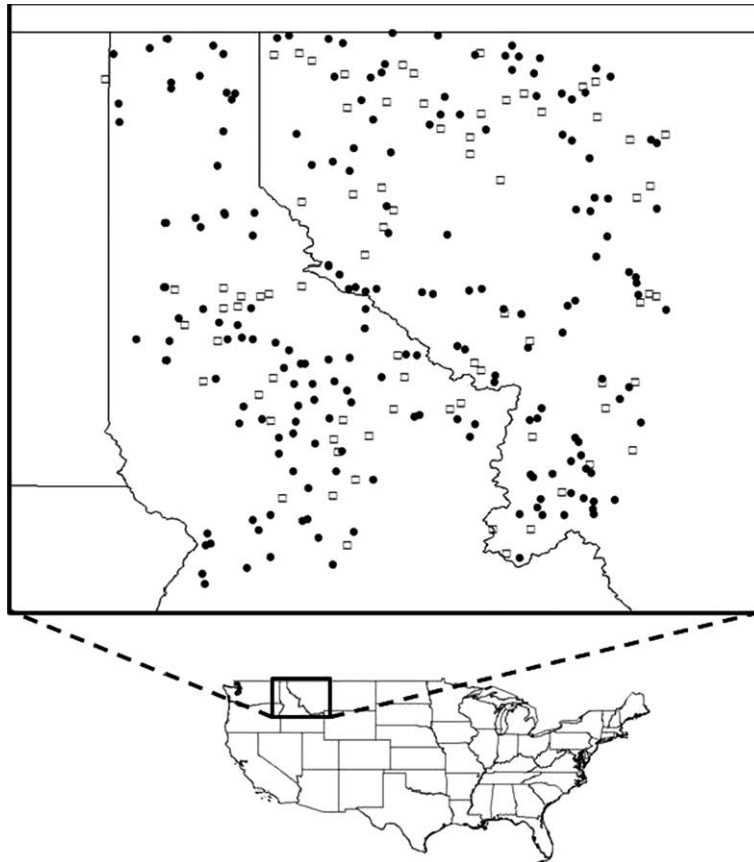


Figure 1. Location of permanently marked Northern Region Landbird Monitoring Program transects used to estimate the distribution of forest birds. The black dots represent transects selected for building species distribution models (SDMs), and the open rectangles represent transects selected for validating SDMs.

### Habitat and detection covariates

We modeled occurrence as a function of several habitat covariates. We selected unique habitat covariates for each species, based on published accounts of habitat use (Table 1). Additionally, we modeled the probability of occurrence as a function of elevation and survey date. Elevation is an indirect gradient that is likely to strongly influence the probability of occurrence (Austin 2002). Survey date is likely to influence the probability of occurrence, especially early in the season as species are arriving from wintering habitat and making settlement decisions. We modeled the probability of occurrence as a function of elevation for all species except varied thrush and chestnut-backed chickadee, because elevation was highly correlated with the presence of mesic forest ( $r = -0.55$ ), a habitat variable used to model the occurrence of both species. Similarly, we modeled the probability of occurrence as a function of date for all species except pileated woodpecker and chestnut-backed chickadee, because they are resident species in western Montana and northern Idaho.

We derived all habitat covariates from Geographic Information System (GIS)-based vegetation measures. All original GIS layers for canopy cover, diameter at breast height (DBH), and land-cover type were 15-m resolution digital land-cover maps developed by the United States Forest Service Northern Region Vegetation Mapping Program (USFS R1-VMP), based on 2001 Landsat TM

imagery and aerial photo interpretation (Brewer et al. 2004). We derived habitat variables from three R1-VMP GIS layers: tree-diameter, canopy cover, and life form. The original R1VMP Tree-diameter layer described tree size with four size categories; we used a Principal Component Analysis (PCA) to reduce the number of DBH variables to two. The original R1VMP Canopy Cover layer described canopy cover using three categories. We again used PCA to reduce the number of canopy cover variables to two. In both cases, one principal component reflected a linear gradient of canopy cover or DBH, whereas the other component reflected a non-linear gradient (high factor loadings on intermediate categories; see also Fletcher et al. 2010). The original R1VMP Life-Form layer describes the relative canopy cover of several vegetative communities in each cell. From this layer, we derived new layers describing the presence/absence of subalpine forest, mesic forest, and shrubby habitat in the surrounding 100 m. We also derived the percentage of land cover in the surrounding 1-km landscape that contained conifer forest from this layer. We derived stream distances from the United States Geological Survey National Hydrography Dataset and elevation from a 30-m resolution Digital Elevation Model. We obtained mean annual precipitation data from the PRISM Climate Group at Oregon State Univ. (PRISM 2006). We calculated road density by merging 2000 Census Data Tiger roads files for Idaho and Montana (Montana Dept of Commerce 2001) with USFS Region 1 road data (USDA

Table 1. Habitat covariates used to model distribution of landbirds in Montana and Idaho, USA.

Species	Habitat covariates	Citation
Pileated woodpecker <i>Dryocopus pileatus</i>	Tree diameter at breast height, canopy cover, distance to nearest stream, elevation*	Bull and Jackson 1995
Brown-headed cowbird <i>Molothrus ater</i>	Canopy cover, road density, elevation*, date†	Lowther 1993
Townsend's solitaire <i>Myadestes townsendi</i>	Canopy cover, percent conifer forest in surrounding 1 km, elevation*, date†	Bowen 1997
Chestnut-backed chickadee <i>Poecile rufescens</i>	Mean annual precipitation, presence of mesic forest	Dahlsten et al. 2002
Golden-crowned kinglet <i>Regulus satrapa</i>	Tree diameter at breast height, distance to nearest stream, presence of subalpine forest, elevation*, date†	Ingold and Galati 1997
MacGillivray's warbler <i>Oporornis tolmiei</i>	Presence of shrubby habitat, distance to nearest stream, elevation*, date†	Pitocchelli 1995
Townsend's warbler <i>Dendroica townsendii</i>	Canopy cover, tree diameter at breast height, elevation*, date†	Wright et al. 1998
Varied thrush <i>Ixoreus naevius</i>	Canopy cover, presence of mesic forest, mean annual precipitation, date†	George 2000
Swainson's thrush <i>Catharus ustulatus</i>	Canopy cover, percent conifer forest in surrounding 1 km, elevation*, date†	Mack and Yong 2000

\*Linear and quadratic responses were explored.

†Linear and quadratic responses were explored. Note that date was not used as a covariate in Maxent models.

Forest Service 2008), and calculating the total road length within a 1km radius from each cell. We aggregated all GIS layers to a common 200-m (40 000 m<sup>2</sup>) resolution prior to analysis to reflect the grain of our sampling unit (100-m radius counts [31 416 m<sup>2</sup>]).

We also modeled detection probability as a function of several covariates. For all species, we modeled detection probability as a function of date of survey, time of survey, wind speed, stream noise and sky cover/precipitation during survey, canopy cover, and DBH. We explored linear and quadratic effects of date of survey and time of survey. Another potential source of heterogeneity in detection probability is the abundance of individuals at a point count (Royle and Nichols 2003). While we do not include this potentially important source of variation in detection probability, modeling detection probability as a function of an index of abundance would not likely change estimates of occurrence in a meaningful way, since all species considered are highly territorial and >75% of detections on a given point-count were of single individuals.

For Bayesian modeling described below, we standardized all non-categorical habitat and detectability covariates to have a mean 0 and standard deviation 1 to improve model convergence.

## Maximum entropy models

We used a maximum entropy machine learning approach to build SDMs for all species. Maximum entropy approaches rely on presence-only data to model the distributions of species, attempting to avoid the problem of imperfect detection by not considering sites during the modeling process where species were not detected. This approach has recently been shown to have high predictive performance relative to other presence-only approaches (Elith et al. 2006). The maximum entropy principle is to approximate a probability distribution that satisfies any known constraints on species distributions, but is closest to maximum entropy (i.e. the approximated probability distribution is as close to uniform as possible). Environmental variables at known presence locations are used to constrain the probability distribution such that the expected value of an environmental variable under the estimated distribution must match its empirical average (Phillips et al. 2006).

We used Maximum Entropy Species Distribution Modeling ver. 3.2.1 (hereafter, Maxent; Phillips and Dudik 2008) to construct maximum entropy models for each species. To construct the Maxent models, we used point locations for recorded species occurrence over the 1994–2008 sampling period as presence-only inputs and GIS maps of species-specific habitat co-variables (Table 1) as the environmental variables. We initially segregated the presence-only points by year, building separate models for each year and estimating mean suitability by averaging across years. We compared this to a single model built with all presence points, regardless of year. These two approaches provided almost identical estimates of relative suitability (Rota 2009), so we only report results from the model built with all presence points, which is the convention with Maxent modeling (Phillips et al. 2006). Additionally, we fit Maxent models that were constrained to estimate either linear-only or linear and quadratic effects for all habitat covariates, because logistic regression and hierarchical occupancy models only included linear and quadratic effects (see below), as well as unconstrained models that allowed for more complex relationships (Phillips and Dudik 2008). Rankings of constrained Maxent models relative to logistic regression and occupancy models was either unchanged or reduced compared to unconstrained Maxent models, so we only report predictive performance of unconstrained Maxent models. For all Maxent models, we used the default parameter settings recommended by Phillips and Dudik (2008).

## Logistic regression models

We used a Bayesian logistic regression approach to model species occurrence, assuming presence or absence at a site is a Bernoulli random variable:

$$z_i \sim \text{Bernoulli}(\psi_i) \quad (1)$$

where  $z_i$  is the latent occupancy state of site  $i$  ( $z_i \in \{0,1\}$ ; site  $i$  is occupied if  $z_i = 1$ , site  $i$  is unoccupied if  $z_i = 0$ ) and  $\psi_i$  is the probability site  $i$  is occupied. Note that the latent occupancy state of each site is assumed to be detected without error with a logistic regression approach. We further modeled  $\psi_i$  as a function of both fixed and random effects:

$$\text{logit}(\psi_i) = \beta_0 + \boldsymbol{\beta}_{\text{cov}} \times \mathbf{x}_i + \tau_t + \gamma_r \quad (2)$$

Here,  $\beta_0$  is an intercept term,  $\boldsymbol{\beta}_{\text{cov}}$  is a vector of regression parameters associated with the covariates,  $\mathbf{x}_i$  is a vector of habitat covariates at site  $i$ ,  $\tau_t$  is a random effect of transect  $t$  and  $\gamma_r$  is a random effect of year  $r$ . By including a random transect effect, we effectively “block” each site by transect, which can be a useful way to deal with potential spatial autocorrelation (Keitt et al. 2002). If spatial autocorrelation was pervasive in the database, we expected it to occur within transects rather than among transects, because the average distance between transects is  $>8$  km. Additionally, we include a random year effect because many sites were surveyed in multiple years. Including a random year effect limits potential pseudoreplication from temporally correlated repeated measures.

## Hierarchical occupancy models

We also used a hierarchical occupancy modeling approach to model species occurrence, assuming presence or absence at a site is a Bernoulli random variable and modeling  $\psi_i$  as a function of both fixed and random effects as above. Note, however, that with this occupancy modeling approach we no longer assume the latent occupancy state is detected without error.

Our hierarchical occupancy model includes an additional observation model to account for imperfect detection. We employed a “removal” sampling protocol (sensu MacKenzie and Royle 2005, Rota et al. 2009), where a species was only surveyed for until it was first detected, for a maximum of two surveys. Such designs are typical of many point-count based surveys (Ralph et al. 1993). We assume detection or non-detection of a species at site  $i$  is dependent on the latent occupancy state at site  $i$ , i.e. we cannot detect a species at site  $i$  if that site is unoccupied. We model detection or non-detection of a species as a Bernoulli random variable:

$$y_i \sim \text{Bernoulli}(1 - (1 - [p_i \times z_i])^J) \quad (3)$$

where  $y_i$  is a binary indicator of whether a species was detected at site  $i$  ( $y_i = 1$ ) or not ( $y_i = 0$ ),  $p_i$  is the probability of detecting a species during an individual survey at site  $i$ , and  $J$  is the maximum number of sampling intervals (note that  $J = 2$ ). The number of surveys conducted is then dependent on whether a species is detected at site  $i$  or not. If a species is detected at site  $i$ :

$$\text{prob}(j_i = k) = p_i(1 - p_i)^{k-1} \quad (4)$$

where  $j_i$  is the sampling interval a species was detected in at site  $i$  and  $k$  is an integer ( $k \in \{1, \dots, J\}$ ) representing the sampling intervals a species could potentially be detected in. If a species is not detected at site  $i$ :

$$\text{prob}(j_i = J) = 1. \quad (5)$$

We further modeled  $p_i$  as a linear function of covariates likely to influence detection probability as:

$$\text{logit}(p_i) = \alpha_0 + \boldsymbol{\alpha}_{\text{cov}} \times \mathbf{v}_i + \rho_o \quad (6)$$

where  $\alpha_0$  is an intercept term,  $\boldsymbol{\alpha}_{\text{cov}}$  is a vector of regression parameters associated with the covariates,  $\mathbf{v}_i$  is a vector of detection covariates at site  $i$ , and  $\rho_o$  is a random effect of observer  $o$ . While our detectability model cannot distinguish between “difficult” and “easy” to detect species (Farnsworth et al. 2002), we are able to explain the bulk of site heterogeneity in detection probability via covariates. We included the potential for random observer effects because similar datasets (e.g. the Breeding Bird Survey) have shown strong effects of observers (Sauer et al. 1994).

## Prior distributions

For all fixed effects in Bayesian models, we specified vague normal prior distributions with mean 0 and variance 1000 (Royle and Dorazio 2008, p. 110). We used a non-informative Uniform(0,1) prior distribution on the intercept parameters  $\alpha_0$  and  $\beta_0$ , which we transformed to the logit scale. We specified all random effects to have a mean zero and a non-informative Uniform(0,10) prior distribution for each standard deviation parameter (Gelman 2006).

We evaluated how sensitive model-averaged fixed effects and posterior model probabilities (see below) were to prior distributions. We ran a subset of models with fixed-effect prior distributions set to mean 0 and variance 10. While posterior distributions of fixed-effect parameters were insensitive to prior specification, estimates of posterior model probabilities (see below) were sensitive to prior specification. Indeed, this is a known issue with Bayesian approaches to model selection (Royle and Dorazio 2008, p. 111). Importantly, however, relative rankings of predictive performance (see below) of models run with more informative priors remained unchanged.

## Model averaging

For Bayesian models, we used the model selection strategy described by Kuo and Mallick (1998; see also Royle and Dorazio 2008). This approach estimates the posterior probability of all possible combinations of fixed effects in a model. This involves fitting an additional inclusion parameter,  $w_\theta$ , for each fixed effect  $\theta$ . Each inclusion parameter is assumed to have a Bernoulli distribution with parameter  $\pi_\theta$ , which is the probability that fixed effect  $\theta$  will be included in the model. We specified vague prior distributions on all inclusion parameters as:

$$w_\theta \sim \text{Bernoulli}(0.5).$$

We then estimated the posterior probability of each model directly from the Markov Chain Monte Carlo (MCMC) output by calculating the proportion of times each combination of fixed effects appeared in the output.

We estimated the regression coefficient of each fixed effect by model averaging over all possible model combinations:

$$\hat{\theta} = \sum_{i=1}^R \delta_i \hat{\theta}_i$$

where  $\delta_i$  is the posterior probability of model  $i$ , and  $\hat{\theta}_i$  is the mean of the posterior distribution of fixed effect  $\theta$  for model  $i$ . Note that if fixed effect  $\theta$  is not included in model  $i$ ,  $\hat{\theta}_i = 0$ . We used WinBUGS ver. 1.4 (Spiegelhalter et al. 2003) for all Bayesian models, which we called remotely from R ver. 2.9 (R Development Core Team 2009) using the package R2WinBUGS (Sturtz et al. 2005). We estimated posterior distributions from two Markov chains, each with 150 000 iterations, discarding the first 50 000 as burn-in and saving every fifth iteration thereafter. We assessed convergence by comparing posterior distributions from each independently run chain. See Supplementary material for sample WinBUGS code for Bayesian models.

### Comparing predictive performance

For occupancy and logistic regression models, we used the fixed effects from the training models to predict the probability of occurrence at validation sites. The problem of imperfect detection is still present in the validation data, and ensuring occupancy model predictions were of the same currency as validation data required adjusting predictions of occurrence for imperfect detection (indeed, failure to adjust occupancy model predictions for imperfect detection led to reduced AUC for eight of nine species, Rota unpubl.). We did this by using the model-averaged fixed-effects from the detectability model to predict detection probability at validation sites. We then multiplied the predicted probability of occurrence by the predicted probability of detecting a species at least once to arrive at a prediction of detection or non-detection at each site.

Rather than providing estimates of habitat coefficients, Maxent provides a predictive map of relative probability of occurrence, which is bounded between 0 and 1 at each cell. To estimate model predictions at validation sites, we overlaid the location of each validation site on this predictive map and extracted the estimated relative probability of occurrence at each site.

We used several approaches to assess how well models predicted where species would be detected. We compared predictive performance by calculating the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) plot, which we calculated using the ROCR package (Sing et al. 2005) in R. AUC is a predominant statistic used for assessing predictive performance in species distribution models (Elith et al. 2006); however, it has received some recent criticism (Austin 2007, Lobo et al. 2008). Thus, we additionally compared predictive performance with the Kappa statistic and explored sources of predictive error by calculating false negative and false positive rates (Fielding and Bell 1997). Kappa, false negative rates, and false positive rates are threshold-dependent performance measures, meaning a threshold must be selected to transform probabilities of occurrence or relative suitability to predictions of presence or absence. To calculate the Kappa statistic

and false positive and false negative rates, we selected a threshold that maximized the sum of specificity and sensitivity (Liu et al. 2005), which is identical to minimizing the sum of false negative and false positive rates.

Finally, to compare between SDM approaches, we calculated relative performance of each approach for predicting distributions of each species. This allowed us to evaluate differences in performance for each SDM approach independent of variation in performance between species. For each species, we calculated relative performance as:

$$\Delta_{q,m} = perf_{q,m} - perf_{min}$$

where  $perf_{q,m}$  is the value of performance measure  $q$  for modeling approach  $m$ ,  $perf_{min}$  is the minimum performance measure  $q$  for modeling approach  $m$ , and  $\Delta_{q,m}$  is the relative value of performance measure  $q$  for modeling approach  $m$ . Note that high  $\Delta$  values for AUC and Kappa indicate high relative predictive performance and a value of 0 indicates the lowest relative predictive performance, while high  $\Delta$  values for False Negative and False Positive rates indicate high relative error rates and a value of 0 indicates the lowest relative error rate.

### Results

Overall, occupancy models tended to have the greatest predictive performance. Occupancy models had the highest AUC scores for seven out of nine species and the highest Kappa scores for five out of nine species. Logistic regression models performed nearly as well as occupancy models, and Maxent consistently had the lowest predictive performance (Table 2, Fig. 2). Accordingly, occupancy models had the highest relative AUC across all nine species (mean  $\Delta_{AUC,O} = 0.075$ , SD = 0.069), followed by logistic regression (mean  $\Delta_{AUC,LR} = 0.057$ , SD = 0.072) and Maxent (mean  $\Delta_{AUC,M} = 0.013$ , SD = 0.024, Fig. 2a). Logistic regression models had the highest relative Kappa (mean  $\Delta_{Kap,LR} = 0.089$ , SD = 0.105), followed by occupancy models (mean  $\Delta_{Kap,O} = 0.084$ , SD = 0.100) and Maxent models (mean  $\Delta_{Kap,M} = 0.010$ , SD = 0.017, Fig. 2b).

False negative and false positive rates also varied among SDM approaches. Relative false negative rates were lowest for Maxent models (mean  $\Delta_{FN,M} = 0.112$ , SD = 0.155), intermediate for occupancy models (mean  $\Delta_{FN,O} = 0.130$ , SD = 0.143) and greatest for logistic regression models (mean  $\Delta_{FN,LR} = 0.163$ , SD = 0.118). Relative false positive rates were lowest for logistic regression models (mean  $\Delta_{FP,LR} = 0.074$ , SD = 0.145), intermediate for occupancy models (mean  $\Delta_{FP,O} = 0.089$ , SD = 0.122) and much higher for Maxent models ( $\Delta_{FP,M} = 0.190$ , SD = 0.184).

Relative predictive performance varied as a function of detection probability when making pair-wise comparisons between models (Fig. 3). Occupancy model AUC scores tended to be high relative to logistic regression models for species with low detection probability, with scores becoming more similar as detection probability increased (Fig. 3a). However, this potential relationship was not as apparent when comparing relative Kappa values. Conversely, occupancy model and logistic regression AUC and Kappa scores tended to be higher relative to Maxent models for

Table 2. Performance of models used to predict the distribution of landbird species in Montana and Idaho, USA.

Species (code)	Model	Detection probability	AUC	Kappa	False negative	False positive
Brown-headed cowbird (BHCO)	Occupancy	0.430	0.697	0.085	0.365	0.306
	Logistic		0.586	0.034	0.346	0.468
	Maxent		0.660	0.046	0.048	0.623
Chestnut-backed chickadee (CBCH)	Occupancy	0.368	0.760	0.143	0.137	0.368
	Logistic		0.752	0.157	0.151	0.340
	Maxent		0.765	0.144	0.144	0.364
Golden-crowned kinglet (GCKI)	Occupancy	0.724	0.651	0.198	0.355	0.397
	Logistic		0.618	0.166	0.311	0.467
	Maxent		0.623	0.143	0.235	0.559
MacGillivray's warbler (MGWA)	Occupancy	0.866	0.652	0.193	0.316	0.434
	Logistic		0.651	0.184	0.319	0.441
	Maxent		0.522	0.042	0.045	0.879
Pileated woodpecker (PIWO)	Occupancy	0.390	0.581	0.056	0.474	0.376
	Logistic		0.557	0.022	0.142	0.754
	Maxent		0.581	0.061	0.532	0.323
Swainson's thrush (SWTH)	Occupancy	0.941	0.795	0.459	0.310	0.231
	Logistic		0.785	0.454	0.348	0.202
	Maxent		0.588	0.137	0.181	0.669
Townsend's solitaire (TOSO)	Occupancy	0.473	0.638	0.072	0.129	0.677
	Logistic		0.670	0.148	0.384	0.355
	Maxent		0.635	0.110	0.453	0.356
Townsend's warbler (TOWA)	Occupancy	0.956	0.738	0.357	0.126	0.511
	Logistic		0.737	0.354	0.231	0.413
	Maxent		0.697	0.292	0.208	0.497
Varied thrush (VATH)	Occupancy	0.868	0.727	0.201	0.224	0.448
	Logistic		0.719	0.288	0.503	0.181
	Maxent		0.610	0.122	0.431	0.390

species with higher detection probability, and this difference declined as detection probability declined (Fig. 3b, c).

False negative and positive rates showed weak variation along a gradient in detection probability (Fig. 4). Logistic regression false negative rates were greatest for highly detectable species, while false positive rates were greatest for difficult to detect species. Maxent models showed the opposite tendency, with the greatest false negative rates occurring for difficult to detect species, and the greatest false positive rates occurring for highly detectable species. This suggests the potential for different tradeoffs in errors associated with these modeling approaches.

## Discussion

While numerous methods exist for modeling the geographic distribution of species, very few explicitly address the problem of detectability. Occupancy models that account for imperfect detection are a promising new method in this regard. Indeed, our hierarchical occupancy models generally demonstrated high predictive performance relative to approaches that did not incorporate imperfect detection, especially for difficult to detect species. However, while occupancy modeling approaches generally led to improved predictions of where species would be detected, the absolute increase in performance measures was relatively small. For example, occupancy model AUC scores for the seven species with the highest relative AUC scores were <2% greater, on

average (range = 0.03–5.55%), than the next highest AUC score.

The small improvement in predictive performance provided by occupancy models was somewhat surprising, but can potentially be explained by three factors: 1) heterogeneity in detection probability, 2) the limited information used in estimating detection probability, and 3) the data used for model validation. The amount of heterogeneity present in estimates of detection probability for each species varied widely, as illuminated by correlations between detectability and covariate(s), and could contribute to variation in performance of occupancy models. For example, a strong negative relationship with detection probability and wind speed would lead to heterogeneity in detection probability, since a species would be more difficult to detect as wind speed increased. Both the brown-headed cowbird and pileated woodpecker had at least one relatively large model-averaged detection probability coefficient, which translates to heterogeneous estimates of detection probability across sites. However, neither the chestnut-backed chickadee nor the Townsend's solitaire had any relatively large model-averaged detection probability coefficient, such that estimates of detection probability were relatively homogeneous across sites. Occupancy models performed better for brown-headed cowbird and pileated woodpecker relative to logistic regression models, while logistic regression models tended to perform better for Townsend's solitaire and chestnut-backed chickadee relative to occupancy models, suggesting that heterogeneity in

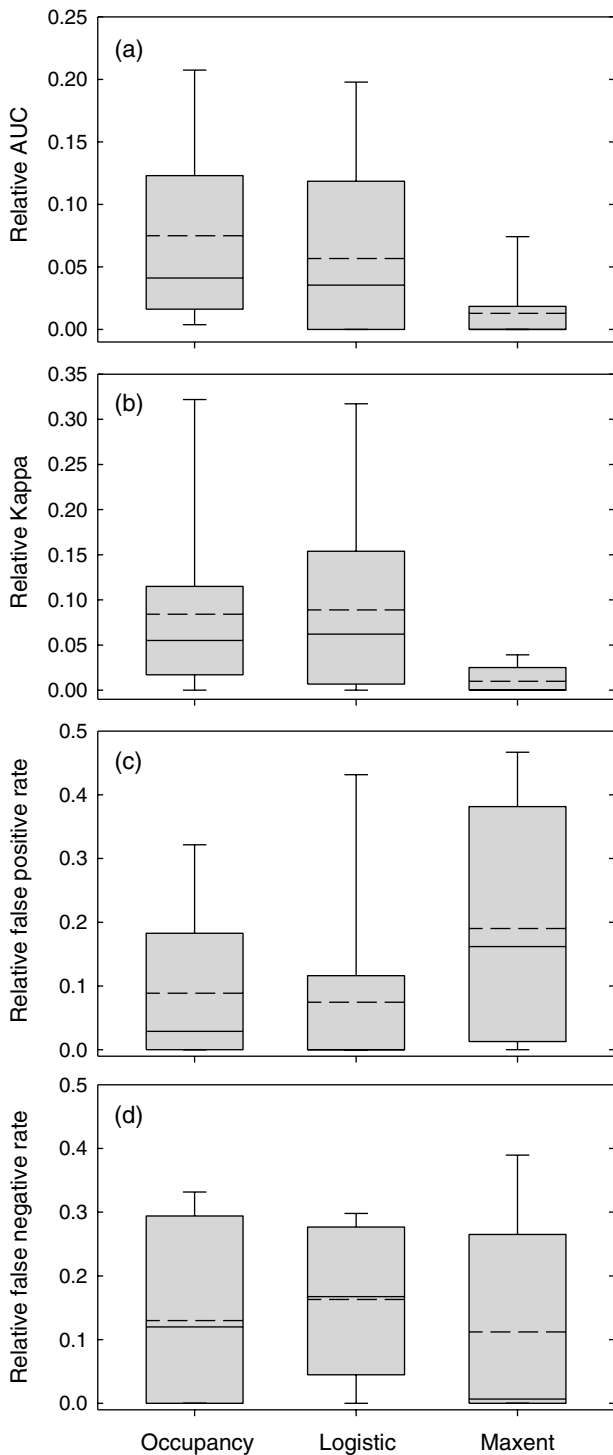


Figure 2. Relative performance of occupancy models, logistic regression models, and Maxent models used to predict the distributions of nine landbird species in Montana and Idaho, USA. Predictive performance measures include (a) area under the curve (AUC) of a receiver operator characteristic plot, (b) Kappa, (c) false positive rate, and (d) false negative rate. A higher value indicates that a given model has a higher rate relative to another model type for a given species. Boundary of boxes indicates the 25th and 75th percentile, solid lines within boxes mark the medians and dotted lines the means, and whiskers indicate the 90th and 10th percentiles.

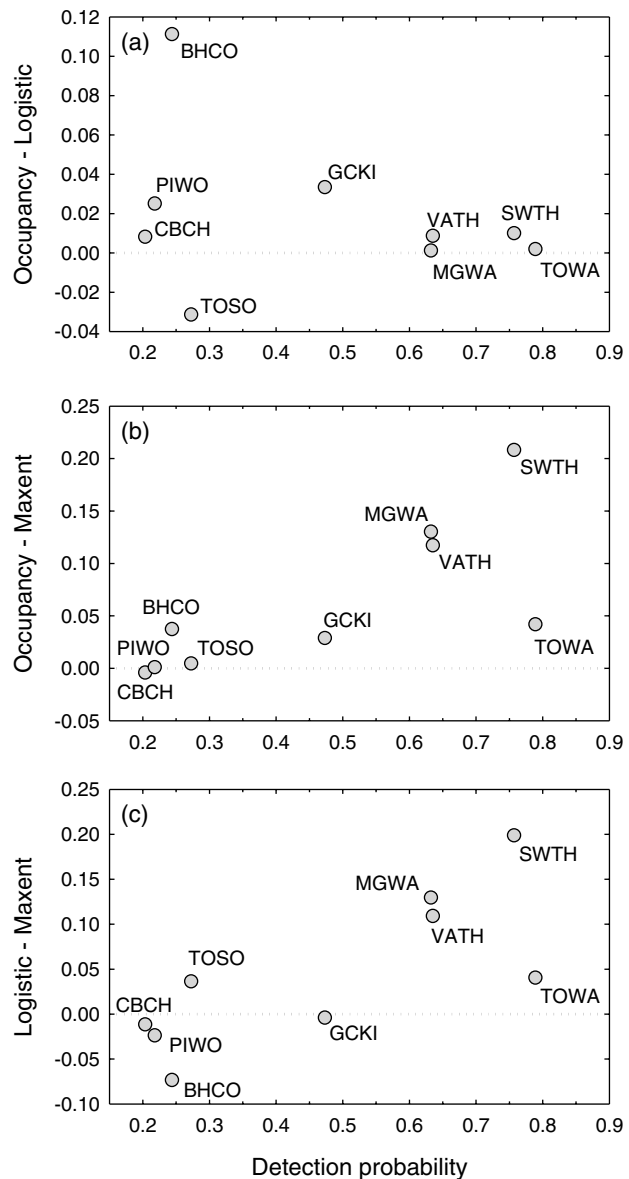


Figure 3. Relative pairwise comparisons of area under the curve (AUC) of receiver operator characteristic plots from models used to estimate the distributions of nine landbird species in Montana and Idaho, USA, as a function of species' detectability. Comparisons are with (a) occupancy and logistic regression models, (b) occupancy and Maxent models, and (c) logistic regression and Maxent models. For description of species codes, see Table 2.

detection probability may lead to increased predictive performance of occupancy models, at least for difficult-to-detect species.

Indeed, simulations demonstrate that for occupancy models to improve predictions of where species will be detected relative to logistic regression models, detection probability must be heterogeneous across sites (Rota unpubl.). Failure to incorporate imperfect detection in logistic regression models will lead to biased estimates of habitat relationships (Tyre et al. 2003, Gu and Swihart 2004). Additionally, if both occupancy and detection



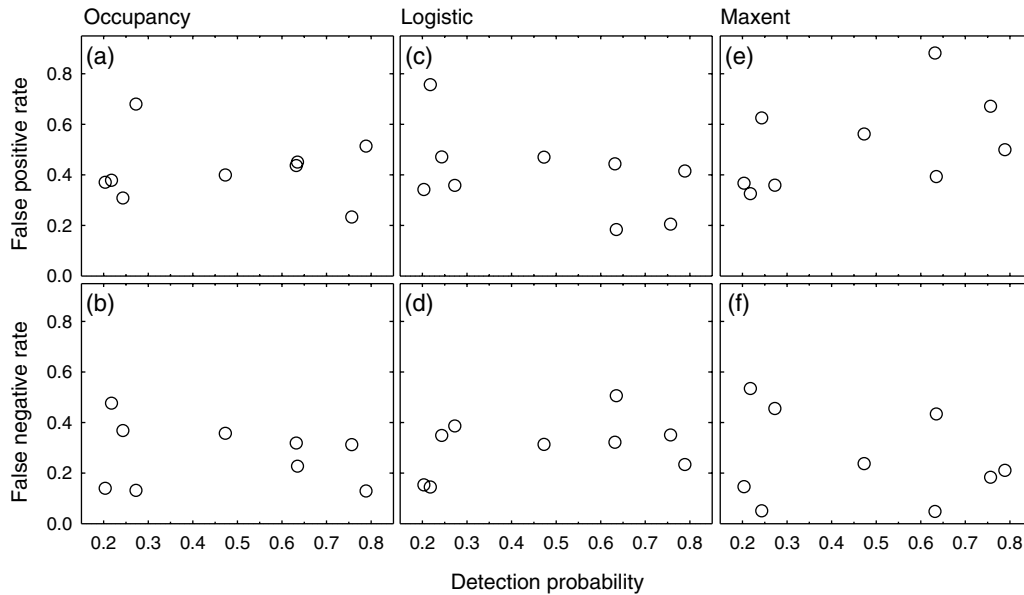


Figure 4. False negative and false positive rates of nine landbird species in Montana and Idaho, USA, as a function of species detectability. False positive and false negative rates as a function of detection probability are shown for (a, b) occupancy models, (c, d) logistic regression models, and (e, f) Maxent models.

probability are correlated with the same covariate (e.g. if both occurrence and detection probability increase along a single habitat gradient) estimates of habitat relationships from logistic regression models can be misleading (MacKenzie et al. 2006). However, if detection probability is homogeneous across sites, relative estimates of habitat relationships will remain the same and occupancy models may not improve predictions of where species will be detected relative to logistic regression models (Rota unpubl.). Therefore, if occupancy models are to be used with the goal of improving predictive performance of SDMs, careful consideration needs to be put into selecting relevant factors affecting detection probability, and variables that are likely to be highly variable between sites may be especially useful.

A second reason for the minimal gain in predictive performance provided by occupancy models may be the short temporal scale of the replicate surveys used for estimating detection probability. Estimates of detection probability were based on two 5-min surveys, which may not provide reliable estimates of detection probability for difficult to detect species (MacKenzie and Royle 2005). Additionally, replicate surveys may have suffered from a lack of independence, since they were conducted in rapid succession. Generally, a tradeoff exists between the number of sites that can be surveyed and the number of replicate surveys conducted at each site (MacKenzie and Royle 2005). Our approach attempted to strike a balance by conducting replicate surveys during each site visit, which had the added benefit of minimizing the time over which closure was assumed for the occupancy models (Rota et al. 2009). The removal modeling approach we employed also makes issues of independence between replicate surveys less problematic, since surveys stop once a species is detected. Despite conducting only two replicate surveys at each site,

however, model averaged estimates of detection probability suggested strong influences of some factors affecting detection probability for six out of nine species. Users of any species distribution modeling approach need to carefully consider the temporal scale of replicate surveys and the complex interplay between how many sites to survey, how many replicate surveys to conduct at each site, the length of replicate surveys, and maximizing closure between surveys.

The short duration of the surveys may not only have made estimating detection probability difficult for some species, but may also have affected the quality of the validation data. Data used for model validation also included the potential problem of uncertain detection and some species almost certainly went undetected at highly suitable validation sites where they were present. Such events may lead to increased false positive rates when calculating predictive performance (i.e. models predicting a species as present, but validation data indicating them as undetected). Indeed, this is an implicit problem that is likely to plague all validation datasets to some degree.

While the incorporation of detection probability improved predictive performance only slightly relative to approaches that ignored detection probability, SDM approaches that included detection/non-detection data consistently outperformed SDM approaches that relied on presence-only data. This result is consistent with earlier work that contrasted other presence-only methods with logistic regression (Brotons et al. 2004). Interestingly, relative performance of Maxent models showed at clear gradient along estimated detection probability. While relative performance of Maxent models was similar to other approaches at lower estimated detection probabilities, Maxent was clearly and consistently inferior to both occupancy and logistic regression models at higher estimated detection probabilities. This could be a result of the

amount of information contained in non-detections for highly detectable vs difficult-to-detect species. For highly detectable species, non-detections may be relatively more informative, whereas for difficult to detect species, non-detections are much more ambiguous, since an individual may be present but undetected. Ignoring non-detection locations for highly detectable species may result in discarding large amounts of informative data and result in the observed poor relative predictive performance of Maxent models for those species. Similarly, Maxent may perform poorly for common species (often positively correlated with detection probability) because randomly selected pseudo-absences are improperly constraining the estimated distributions of those species. Regardless of the cause, this result suggests that Maxent modeling approaches may be adequate for modeling distributions of rare and elusive species, but that detection/non-detection data should be collected whenever possible.

The poor predictive performance of Maxent models was due to relatively high false positive rates. This may stem from the predictive nature of these models. Occupancy and logistic regression models attempt to predict probability of occurrence, while Maxent models predict a relative measure of occurrence/suitability (Phillips et al. 2006). The probabilistic nature of occupancy and logistic regression model predictions accounts for prevalence of a species, such that habitat predicted as high quality may still have a low probability of occurrence because a species is generally rare across the landscape. Because Maxent only considers relative suitability when making predictions, it can predict a high relative suitability in high quality habitat, even for relatively rare species.

The source of predictive errors can give important insight into the utility of SDMs in a conservation context, since some types of error may be more acceptable than others. False positives are potentially more acceptable than false negatives because they are more easily explained by sampling and biological processes. For example, a species may be present in suitable habitat but remain undetected. Additionally, metapopulation theory predicts that suitable habitat may be unoccupied through stochastic processes alone. However, false negative rates may pose more problematic biological or predictive explanations. For example, if a species is present in habitat predicted as unsuitable, that species may be occupying an ecological trap, or models may be incorrectly predicting suitability. Both logistic regression and Maxent models showed a weak relationship with false positive rates along a detectability gradient, with logistic regression showing the highest false positive rates at low detection probability, and Maxent showing the highest false positive rates at high detection probability. This suggests that logistic regression may be most useful with highly detectable species, while Maxent may be most useful with difficult to detect species. However, occupancy models showed no strong relationship with detection probability and any source of predictive error, suggesting this approach can perform as well for highly detectable species as for difficult to detect species.

Despite the inability of occupancy models to dramatically improve predictive performance of species distribution models, there are several advantages to using such an

approach. For example, in situations where detection probability is relatively homogeneous across sites, use of occupancy models is unlikely to result in improved predictive performance. However, in these situations, occupancy models still should provide unbiased estimates of habitat relationships and the probability of occurrence. Additionally, predictions from occupancy modeling approaches are more directly interpretable than other SDM approaches, since estimates of occurrence are distinct from estimates of detection probability. Hierarchical occupancy models can also address common problems associated with SDMs, such as spatial and temporal autocorrelation in data (Dormann et al. 2007). While further information must be collected in datasets to allow estimating detection probability, our example shows that such information can be collected at little to no additional cost by simply collecting ancillary information during surveys (e.g. time-to-first detection).

We have demonstrated that occupancy modeling approaches hold promise for improving models of species distributions. However, results shown here suggest that occupancy models do not always substantially change predictive performance of species distribution models, and other recent work suggests they may not affect inference regarding some behavioral and ecological questions in some situations (Fletcher 2009). In these instances, issues of sampling design other than imperfect detection may be of greater importance for inference (Bart et al. 2004). Ecologists must first consider the exact biological interpretation of a species' geographic distribution they wish to acquire and design data collection accordingly. Then, regardless of the sampling design or modeling approach, the onus is on ecologists to address the detection issue and other problems in sampling design in rigorous ways to evaluate potential biases in inference. This application highlights that practitioners need to balance the additional effort of an occupancy modeling approach relative to potential gains in improved predictions of species distributions.

*Acknowledgements* – This work was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant #2006-55101-17158. The landbird database was created through support from USFS Northern Region (03-CR-11015600-019). We thank J. Young and A. Nosen for providing support on the avian database and GIS. We thank B. Dorazio for advice on hierarchical modeling and M. Araújo, D. MacKenzie, R. Gitzen, W. Thogmartin, and an anonymous reviewer for helpful comments on earlier versions of this manuscript.

## References

- Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. – *Ecol. Model.* 200: 1–19.
- Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. – *Ecol. Model.* 157: 101–118.
- Bart, J. et al. 2004. Goals and strategies for estimating trends in landbird abundance. – *J. Wildl. Manage.* 68: 611–626.

- Bowen, R. V. 1997. Townsend's solitaire (*Myadestes townsendi*). – In: Poole, A. (ed.), The birds of North America online. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds.cornell.edu/bna/species/269>>.
- Brewer, C. K. et al. 2004. Northern Region Vegetation Mapping Project Summary Report and Spatial Datasets. – Northern Region USFS.
- Brotons, L. et al. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. – *Ecography* 27: 437–448.
- Bull, E. L. and Jackson, J. A. 1995. Pileated woodpecker (*Dryocopus pileatus*). – In: Poole, A. (ed.), The birds of North America online. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds.cornell.edu/bna/species/148>>.
- Carroll, C. and Johnson, D. S. 2008. The importance of being spatial (and reserved): assessing northern spotted owl habitat relationships with hierarchical Bayesian models. – *Conserv. Biol.* 22: 1026–1036.
- Dahlsten, D. L. et al. 2002. Chestnut-backed chickadee (*Poecile rufescens*). – In: Poole, A. (ed.), The birds of North America online. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds.cornell.edu/bna/species/689>>.
- Dormann, C. F. et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. – *Ecography* 30: 609–628.
- Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.
- Farnsworth, G. L. et al. 2002. A removal model for estimating detection probabilities from point-count surveys. – *Auk* 119: 414–425.
- Fernandez, N. et al. 2006. Landscape evaluation in conservation: molecular sampling and habitat modeling for the Iberian lynx. – *Ecol. Appl.* 16: 1037–1049.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – *Environ. Conserv.* 24: 38–49.
- Fletcher, R. J. Jr 2009. Does attraction to conspecifics explain the patch-size effect? An experimental test. – *Oikos* 118: 1139–1147.
- Fletcher, R. J. Jr et al. 2010. Insights from ecological theory on temporal dynamics and species distribution modeling. – In: Drew, A. et al. (eds), Predictive modeling in landscape ecology. Springer, in press.
- Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. – *Bayesian Analysis* 1: 515–533.
- George, T. L. 2000. Varied thrush (*Ixoreus naevius*). – In: Poole, A. (ed.), The birds of North America online. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds.cornell.edu/bna/species/541>>.
- Gu, W. D. and Swihart, R. K. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. – *Biol. Conserv.* 116: 195–203.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – *Ecol. Model.* 135: 147–186.
- Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – *Ecol. Lett.* 8: 993–1009.
- Hutto, R. L. and Young, J. S. 1999. Habitat relationships of landbirds in the northern region. – USDA Forest Service, general technical report RMRS-GTR-32, USDA Forest Service.
- Hutto, R. L. and Young, J. S. 2002. Regional landbird monitoring: perspectives from the Northern Rocky Mountains. – *Wildl. Soc. Bull.* 30: 738–750.
- Ingold, J. L. and Galati, R. 1997. Golden-crowned kinglet (*Regulus satrapa*). – In: Poole, A. (ed.), The birds of North America online. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds/cornell.edu/bna/species/301>>.
- Keitt, T. H. et al. 2002. Accounting for spatial pattern when modeling organism-environment interactions. – *Ecography* 25: 616–625.
- Kuo, L. and Mallick, B. 1998. Variable selection for regression models. – *Indian J. Stat.* 60B: 65–81.
- Lassalle, G. et al. 2008. Diadromous fish conservation plans need to consider global warming issues: an approach using biogeographical models. – *Biol. Conserv.* 141: 1105–1118.
- Liu, C. et al. 2005. Selecting thresholds of occurrence in the prediction of species distributions. – *Ecography* 28: 385–393.
- Lobo, J. M. et al. 2008. AUC: a misleading measure of the performance of predictive distribution models. – *Global Ecol. Biogeogr.* 17: 145–151.
- Lowther, P. E. 1993. Brown-headed cowbird (*Molothrus ater*). – In: Poole, A. (ed.), The birds of North America online. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds.cornell.edu/bna/species/047>>.
- Mack, D. E. and Yong, W. 2000. Swainson's thrush (*Catharus ustulatus*). – In: Poole, A. (ed.), The birds of North America online. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds.cornell.edu/bna/species/540>>.
- MacKenzie, D. I. and Royle, J. A. 2005. Designing occupancy studies: general advice and allocating survey effort. – *J. Appl. Ecol.* 42: 1105–1114.
- MacKenzie, D. I. et al. 2002. Estimating site occupancy rates when detection probabilities are less than one. – *Ecology* 83: 2248–2255.
- MacKenzie, D. I. et al. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. – Elsevier.
- Manly, B. F. J. et al. 2002. Resource selection by animals: statistical design and analysis for field studies, 2nd ed. – Kluwer.
- Marsh, D. M. and Trenham, P. C. 2008. Current trends in plant and animal population monitoring. – *Conserv. Biol.* 22: 647–655.
- Martin, T. G. et al. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. – *Ecol. Lett.* 8: 1235–1246.
- Montana Dept of Commerce 2001. TIGER redistricting 2000 data, with 2000 census block population counts. – Montana census and economic information center, <<http://nris.mt.gov/nsdi/nris/tgr2000r.html>>.
- Phillips, S. J. and Dudik, M. 2008. Modeling of species distribution with Maxent: new extensions and a comprehensive evaluation. – *Ecography* 31: 161–175.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Phillips, S. J. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – *Ecol. Appl.* 19: 181–197.
- Pitocchelli, J. 1995. MacGillivray's warbler (*Oporornis tolmiei*). – In: Poole, A. (ed.), The birds of North America online. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds.cornell.edu/bna/species/159>>.
- PRISM 2006. Precipitation: annual climatology (1971–2000). – PRISM Climate Group, Oregon State Univ., <<http://prism.oregonstate.edu>>, accessed 30 June, 2009.
- R Development Core Team 2009. R: a language and environment for statistical computing. – R Foundation for Statistical Computing, Vienna, Austria, <[www.R-project.org](http://www.R-project.org)>.
- Ralph, C. J. et al. 1993. Handbook of field methods for monitoring landbirds. – General technical report PSW-GTR-144, USFS Pacific Southwest Research Station.
- Rota, C. T. 2009. Confronting assumptions of occupancy estimation and species distribution models: the problem of detectability in wildlife populations. – MS thesis, The Univ. of Florida, Gainesville, FL.

- Rota, C. T. et al. 2009. Occupancy estimation and the closure assumption. – *J. Appl. Ecol.* 46: 1173–1181.
- Royle, J. A. and Nichols, J. D. 2003. Estimating abundance from repeated presence–absence data or point counts. – *Ecology* 84: 777–790.
- Royle, J. A. and Kéry, M. 2007. A Bayesian state-space formulation of dynamic occupancy models. – *Ecology* 88: 1813–1823.
- Royle, J. A. and Dorazio, R. M. 2008. Hierarchical modeling and inference in ecology. – Academic Press.
- Sauer, J. R. et al. 1994. Observer differences in the North-American breeding bird survey. – *Auk* 111: 50–62.
- Sing, T. et al. 2005. ROCR: visualizing classifier performance in R. – *Bioinformatics* 21: 3940–3941.
- Spiegelhalter, D. J. et al. 2003. WinBUGS user manual, version 1.4. – MRC Biostatistics Unit, Cambridge, <[www.mrc-bav.com.ac.uk/bugs/](http://www.mrc-bav.com.ac.uk/bugs/)>.
- Sturtz, U. et al. 2005. R2WinBUGS: a package for running WinBUGS from R. – *J. Stat. Software* 12: 1–16.
- Thogmartin, W. E. et al. 2004. A hierarchical spatial model of avian abundance with application to cerulean warblers. – *Ecol. Appl.* 14: 1766–1779.
- Tyre, A. J. et al. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. – *Ecol. Appl.* 13: 1790–1801.
- USDA Forest Service 2008. Travel routes for region 1. – USDA Forest Service, R1, regional office engineering, <[www.fs.fed.us/rl/gis/thematic\\_data/TravelRoutesR1.htm](http://www.fs.fed.us/rl/gis/thematic_data/TravelRoutesR1.htm)>.
- Ward, G. et al. 2008. Presence-only data and the EM algorithm. – *Biometrics* 65: 554–563.
- Wright, A. L. et al. 1998. Townsend's warbler (*Dendroica townsendi*). – In: Poole, A. (ed.), *The birds of North America online*. Cornell Lab of Ornithology, Ithaca, <<http://bna.birds.cornell.edu/bna/species/333>>.

Download the Supplementary material as file E6433 from <[www.oikos.ekol.lu.se/appendix](http://www.oikos.ekol.lu.se/appendix)>.